Mouse *Sebox* homeobox gene expression in skin, brain, oocytes, and two-cell embryos

Mario Cinquanta*†, Alessandra C. Rovescalli*, Christine A. Kozak‡, and Marshall Nirenberg*§

*Laboratory of Biochemical Genetics, National Heart, Lung and Blood Institute, and ‡Laboratory of Molecular Microbiology, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD 20892

Sebox is a mouse paired-like homeobox gene, previously named OG-9. Sebox genomic DNA and cDNA were cloned and sequenced. In addition, rat and human Sebox genomic DNAs were cloned and sequenced, and the predicted amino acid sequences were compared. The mouse Sebox gene was mapped to chromosome 11 near the Evi 2 locus. The mouse Sebox gene is expressed in brain, skin, ovary, and liver of mice. In the brain, the Sebox gene is expressed in the cerebral cortex and CA areas of the hippocampus, pontine nuclei, choroid plexus, and the cerebellum. Northern analysis and RNase protection assays revealed low levels of Sebox RNA in 12-day mouse embryos and higher levels in 18- and 19-day embryos. In late embryos and newborn mice, Sebox expression is localized in the epidermis. In adult mice, Sebox RNA was found in maturing oocytes and in fertilized eggs; however, the abundance of Sebox RNA is decreased in the two-cell embryo, and little or none was detected in the four-cell embryo. Hence, Sebox is a maternally expressed homeobox gene.

transcription factors | OG-9 | rat *Sebox* | human *Sebox* | epidermis

Homeodomain proteins play a key role in coordination of gene activity and determination of cell fate in the development of organisms as diverse as plants, yeast, insects, and mammals (1). The homeodomain region of the proteins, and sometimes the functions of the proteins, have been conserved during evolution. The prd-like class of homeobox genes includes several genes that encode proteins with a homeodomain similar to that of prd but lack a paired domain, which is present in homeodomain proteins of the paired class (2). Most of the prd-like homeodomains have a glutamine residue at position 50 of the homeodomain, whereas a serine residue is found at the same position in the paired homeodomain (2). In addition, most of the prd-like homeobox genes have an intron between codons for amino acid residues 46 and 47 of the homeodomain (2).

Sebox, a *s*kin-*e*mbryo-*b*rain-*o*ocyte-specific homeobox gene (formerly termed *OG-9*), is a prd-like class gene that was cloned previously in this laboratory (3). The amino acid sequence of the Sebox homeodomain is not closely related to any known homeodomain. The murine homeodomain with the highest similarity to Sebox is Phox2 (65% identical amino acid residues) (4).

Here we describe the predicted amino acid sequence of the Sebox protein, the structure and chromosomal location of the *Sebox* gene, and the pattern of expression of the *Sebox* gene, and demonstrate that *Sebox* is a maternally expressed gene. We also identified the human *SEBOX* gene and cloned and sequenced rat *Sebox* genomic DNA.

Materials and Methods

Cloning. The 3′-end of the mouse *Sebox* gene was cloned by rapid amplification of cDNA ends (RACE) by using the 3′-RACE System (Life Technologies, Rockville, MD) and 0.5 μg of poly(A)$^+$ RNA from brains of adult BALB/c mice. A cDNA fragment then was amplified by two sequential PCRs with nested primer pairs (for the first PCR, +P was 5′-TCCGGATCCGAGGAAGAGGAC-CACTTTCAGTGT-3′, and −P was 5′-GGCCACGCGTCGA-

CTAGTAC-3; in the second PCR, +P was 5′-GCGGGATCC-TGGAGCTGGAGCGGGTATT-3′, and −P was as above). The resulting PCR product was subcloned in pBluescriptII-KS(+) (Stratagene) and sequenced. The 5′ termini of *Sebox* mRNAs were identified by RNA ligase-mediated 5′-RACE, according to Schaefer (5). Decapped poly(A)$^+$ RNA was ligated to the 3′ end of 5′-UAAUACGACUCACUAUAGGGAGGAGGAAGCU-CUAUUGCACUAGUCUCGAGGAUCCAAUAAA-3′, and 1 μg of the ligated RNA was used for reverse transcription with a gene-specific primer (GSP1: 5′-GCTTCAGGCAGGTGAGTTA-3′) catalyzed by SuperScript II (Life Technologies). The first-strand cDNA then was used as a template for PCR amplification by using a reverse primer nested to GSP1 (GSP2: 5′-AGGATAGGGC-CTAGCTGCAAATAC-3′) and a forward primer specific for the anchor oligonucleotide (anchor-1: 5′-GGAGGAGGAAGCTC-TATTGCACTA-3′). An additional amplification then was performed by using an aliquot from the previous PCR and primer pair: GSP3 (5′-CACCAACTGCCCGACACTGAAAG-3′; nested to GSP2), and anchor-2 primer (5′-TTGCACTAGTCTCGAG-GATCCAAT-3′; nested to anchor-1). The PCR products were cloned into pCRII (TA Cloning System, Invitrogen) and sequenced in both directions.

Human genomic DNA from placentas of multiple individuals was obtained from CLONTECH and amplified by PCR by using Platinum *Taq* DNA Polymerase High Fidelity (Life Technologies), using two primer pairs: +P: 5′-GAAGTATAGGATT-TCAGGCACAGA-3′; −P: 5′-TGGAGCTGGCAAAGGGTA-AGGT-3′; and +P: 5′-GCCAGCCACCCTGAGCATGATA-3′; −P: 5′-CAGGGAATCCACTTCTGATGTGTT-3′. Deoxyadenylic acid residues then were added to the 3′ termini of amplified DNA products catalyzed by *Taq* DNA polymerase, and DNA fragments were ligated into pCRII. Three to five recombinant clones from each PCR reaction mixture were isolated, and DNA inserts were sequenced in both directions by using M13(-20) and reverse primers.

Similarly, rat genomic DNA from kidneys of 200 outbred Sprague–Dawley male rats was purchased from CLONTECH and amplified with primers based on the mouse *Sebox* genomic sequence (primer pair 1: +P, 5′-GGAACCTTCCCTGGCTCTAG-GATG-3′; −P, 5′-ATAGGATCCCCCAACCCACTGGCACAG-3′; pair 2: +P, 5′-GCCGGATCCCAGTCCTGTGGAGGCATCT-3′; −P, 5′-CTTGGATCCGATGGCTGAGGTATAGATA-AGGTC-3′; pair 3: +P, 5′-TGGGTCCAGGGCAGAGTT-3′; −P, 5′-TTGGGATCCATCACTCTCTCTTCAGATCCACAG-3′).

DNA Sequencing. Cloned DNA fragments were sequenced with a Perkin–Elmer/Applied Biosystems 373A DNA sequencer by using fluorescent dideoxynucleotides. Sequence analysis and assembly of contigs were performed by using the WISCONSIN SEQUENCE ANALYSIS PACKAGE (Genetics Computer Group, Madison, WI).

***Sebox* Chromosomal Mapping.** *Sebox* was mapped by analysis of the cross (NFS/N × *Mus spretus*) × *M. spretus* or C58/J (6).

RNA Isolation and Northern Blot Hybridization Analysis. Total RNA was isolated by a modified version of the Chomczynski and Sacchi method (7) (Totally RNA system, Ambion, Austin, TX), and poly(A)$^+$ RNA was purified by chromatography on oligo-dT cellulose (5 Prime→3 Prime). Northern blots containing 2 μg of poly(A)$^+$ RNA per lane were hybridized with a double-stranded DNA probe spanning the coding region downstream of the homeobox (nucleotide residues 991-1410 in Fig. 2) labeled with [α-^{32}P]dCTP (Random Primed DNA labeling kit, Stratagene). Blots were washed at high stringency (50°C, 0.1 × SSC/0.1% SDS), and autoradiograms were developed after 10–15 days.

***In Situ* Hybridization.** Superovulation was induced in 4-week-old BALB/c mice by injection of pregnant mare serum (PMS, Sigma), 5 units i.p., at 2:00 p.m., followed by human chorionic gonadotrophin (hCG, Sigma), 5 units, 46–48 h later (8). Superovulating mice then were either killed 21 h after hCG to obtain oviducts containing ovulated eggs or mated and killed 24 h after hCG to obtain fertilized eggs. For the analysis of two- and four-cell embryos, oviducts of superovulated or mated females were collected 48 and 68 h after hCG injection, respectively. Ovaries and fallopian tubes containing either unfertilized eggs or one-, two-, or four-cell embryos were fixed in 2% paraformaldehyde (PFA) in PBS for 7–8 h at 4°C, infiltrated with 20% sucrose overnight at 4°C, and frozen in M-1 embedding matrix (Lipshaw Manufacturing, Detroit) over a mixture of 2-methylbutane/dry ice. Other tissues used for *in situ* hybridization were frozen immediately after dissection without fixation. Sections (12 μm) were cut with a cryostat, fixed in 4% PFA, treated with 0.25% acetic anhydride, and dehydrated by passage through a graded ethanol/water series. Unless otherwise stated, a [^{33}P]UTP-labeled RNA probe corresponding to the portion of mRNA encoding the C-terminal part of the protein (nucleotide residues 991-1410 in Fig. 2) was used. Hybridization and washes were performed as described elsewhere (8).

Results and Discussion

Northern Analysis. Northern analysis of adult mouse tissues showed that the *Sebox* gene is expressed in ovary, skin (Fig. 1*B*), and brain, and, to a lesser extent, in liver (Fig. 1*A*). Two diffuse bands, about 1.5 and 1.8 kb in length, were found in adult brain RNA (Fig. 1 *C* and *D*), whereas only one band, approximately 1.45 kb in length, was detected in ovary, skin, and liver (Fig. 1 *A* and *B*). The temporal pattern of expression in postnatal brain and skin was determined in newborn (P0) through adult male mice (Fig. 1*C*). The maximum amount of *Sebox* RNA in skin was achieved by the second postnatal day and was maintained throughout postnatal development and in the adult (Fig. 1*B*). The two brain-specific species of *Sebox* mRNA barely were detected in P0 brain, but both species of RNA gradually increased in abundance during postnatal development, reaching the highest level in adult mice (Fig. 1*C*). This observation suggests a role for *Sebox* in the final differentiation of the brain or maintenance of brain function rather than in cell type specification. *Sebox* expression was analyzed in mid- and late-stage mouse embryos (Fig. 1*D*). The earliest expression of *Sebox* RNA was detected in poly(A)$^+$ RNA from 12.5-day-old mouse embryos; the level of *Sebox* RNA gradually increases during later
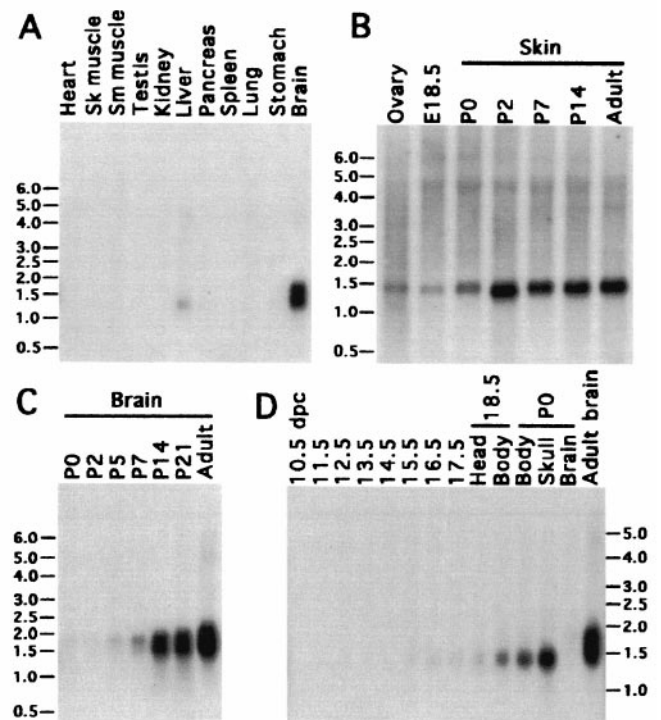
Fig. 1. Northern analysis of *Sebox* mRNA in mouse tissues. Blots (2 μg of poly(A)$^+$ RNA/lane) were hybridized with a cDNA probe labeled with [α-^{32}P]dCTP (specific activity of 1.0 × 10^9 cpm/μg). A 1.45-kb RNA transcript was detected in adult ovary, skin, and liver (*A* and *B*). In addition, RNA of the same length was observed during late fetal development (*D*). Two diffuse bands of RNA, approximately 1.5 and 1.8 kb in length, were detected in postnatal brain starting from the fifth postnatal day, which increased in abundance thereafter (*C*).

development. At 17.5 days, the hybridization signal is a small fraction of the signal detected in the body and head (devoid of the brain) of newborn mice (P0). Expression of *Sebox* RNA in 12.5-day-old embryos also was detected by reverse transcription–PCR (data not shown).

Structure of the Mouse *Sebox* Gene and Cloning of Rat and Human *Sebox* Homologs. A *Sebox* cDNA fragment spanning the region from the beginning of the homeobox to the putative stop codon was isolated by reverse transcription–PCR from adult mouse brain and midstage embryo poly(A)$^+$ RNA. Comparison of the nucleotide sequence of the cDNA fragment and the sequence of *Sebox* genomic DNA (3) revealed an intron between codons for amino acid residues 46 and 47 of the homeodomain. The 5′ and 3′ ends of *Sebox* cDNA were identified by RNA ligase mediated-RACE and 3′ RACE, respectively, by using RNA from either adult brain or ovaries. The *Sebox* gene contains three exons and encodes a putative homeodomain protein, 190 aa residues in length. The ATG encoding the first methionine of the predicted Sebox protein is located within exon 1; part of the homeobox is found in exon 2 and the rest in exon 3. Unlike most homeodomain proteins, the Sebox homeodomain is near the N terminus of the protein rather than the C terminus. The predicted Sebox protein is rich in serine and proline residues, a feature common to the transactivation domains of other transcription factors (9). Excluding the homeodomain, serine and proline residues account for 30% of the total amino acids. Two alternative polyadenylation sites (nucleotides residues G2339 and T2507 shown in Fig. 2) were found by sequencing several clones isolated by 3′ RACE, and two polyadenylation signals were found in the 3′

```
501   GGGCAAGCTGGTGGTGCCATCTGCTCCCTAGCCCTGCTGCCAGGAGGTGGATCCAACGGC
561   ATCTTCATCCTATCCGATCTCAGAGAGCGGGGCAGGAACCCATCTTGTTTCCCCTGAAGG
621   AAGCCAAGAGCAGTGTACTTGGGGATTTTAGGGCACAGCTGAAGCCCTGGGTGTCTGCAG
681   TTTGAGGGAAGAATTTCACAAAGTACAGGATTGAAAGCAGAGACTGTCAGAGGCTGAGGT
741   TTGAGGAGGGAGGTGTGACCAGCTTCTTGGTGCACACCAATCTGGTTGAACTCAGATCTC
801   ATCAACCCCAGTCCTGCCATGTTTCCTACTTCTCAACCTCTCTTGAATGAATGCCAAGCC
861   TCTTCCCTCCAGAGGCGGAGCCACCCACTCTGAGCGGCTTAAAACCCTAAAGGCCTGAGTC
921   TGGTGCGGGGAACAGCCCTGGCATGGCCCTCTGGCTCCTCACCCCATGTCAGCTCCATG
  1                                                              M
981   GCCAGTCCTGTGGAGGCATCTCCAGgtgacctggtggggaaagggtagcaaggtgggctg
  2   A  S  P  V  E  A  S  P  G
1041  gggaagggtggaggatagggaagggtggaagggagagagtagaggttgcagaggctgatc
1101  aaggcttgggggctgggtggaggctcttgcagggagtcttggtttgattcccatcccggc
1161  ccacagGCTGTGCCAGTGGGTTGGGTCCCCATCGGAGGAAGAGGACCACTTTCAGTGTCG
  11        C  A  S  G  L  G  P  H  R  R  K  R  T  T  F  S  V
1221  GGCAGTTGGTGGAGCTGGAGCGGGTATTTGCAGCTAGGCCCCTATCCTGACATCAGCACCC
  29   Q  L  V  E  L  E  R  V  F  A  A  R  P  Y  P  D  I  S  T  R
1281  GTGAGCACCTGGCTCAGGTAACTCACCTGCCTGAAGCCAAGATCCAGgtgagtcacaaga
  49   E  H  L  A  Q  V  T  H  L  P  E  A  K  I  Q
1341  gccaacttctgggctcaacagcctcttctctctgggaattctgctcttgggttttctgtc
1401  ccacaaggctgtgctggtgccaaaatccatccttcatcttttctatcagGTGTGGTTCC
  64                                                    V  W  F  Q
1461  AGAATCGGCGAGCCAAGAGAATCAAGGACAGAAAGCCAGGAGCCCTAAACTCCAGGCTGG
  68   N  R  R  A  K  R  I  K  D  R  K  P  G  A  L  N  S  R  L  E
1521  AGCTTCCCCAAACTCCTGTTCTCTTCCAGACAGTTCAGCTCAGTATACTTCAGCATGCCCACCAC
  88   L  P  P  N  S  C  L  P  D  T  P  Q  L  P  W  D  P  G  T
1581  CATCAAGCCATCCTCTGCACCCTACCAGTTCAGCTCAGTATACTTCAGCATGCCCACCAC
  108  S  S  H  P  L  H  P  T  S  S  A  Q  Y  T  S  A  C  P  P  Q
1641  AGACCTCCTGGCTCAGGCCCCATTTTGGGTCCAGGGCAGAGTTGGTCAGGGGCTAAAGTGG
  128  T  S  C  L  G  P  I  L  G  P  G  Q  S  W  S  G  A  K  V  A
1701  CAGCCCCATGGGGGACAAGTGGGGCTTCAGGGATTCACTCTTCTTTAGAGCAAATTGTTC
  148  A  P  W  G  T  S  G  A  S  G  I  H  S  S  L  E  Q  I  V  P
1761  CTCAGACTTCACTGGGCAACCTGTCTGACCTTATCTATACCTCAGCCATCGTCACCAATG
  168  Q  T  S  L  G  N  L  S  D  L  I  Y  T  S  A  I  V  T  N  V
1821  TAGACCACTTCTAATTTAGGTGTAGAGCTTTTAAGTCTTGGCTCTGGCCTTGTAGCCCAT
  188  D  H  F  *  190
1881  AGGACAGAGCACATAAGAGAACTCCTGCCTCTTTTTTTGCACAGAGTGTTTGCTAACAGT
1941  TTGGCTCAGGGACTCCCACCCCCATCCCTAGTCTCCACCAGTTTAGTTGGGGTCCTGTGA
2001  CTGGAGCCTGCCTCCTAAAACTCGGCATCCTGGTCCTTTTGGGGACATGAGCATCAGGAG
2061  GTACTGTTAAAGTGAGGGACTACATTTTCCACCTTTACTGACAGGTTTGGGTCCCCAACT
2121  CCTATTATCTTAGCAGAATGATCTCCTTCCTCTAAAGGGGCAGCCACCAGTCTAAGAGGG
2181  CCCTTCAGGACTCACATATGTGCCAGGGGCAGGGTACTGGTCTGCTGGAGCCCACATTT
2241  TCAAGGTTCAGGGCCCCATTGCCTGACTTCCTGCCACCTGGACAAATTAATCAGAACTGT
2301  GGATCTGAAGAGGAGGTGATATTAAAGTTACAGAAATAGAAGCCTGGGTCTTAGCTTTAT
2361  TGAATTTCCCCAAAGATCAATTGGCTGCCTTCTGATGTGGGCTGCGGGGACAACCTAAC
2421  AGAACTGGGGGGTGGGGGTGGGGTGTGTGGTATCACTGGCATGTGTAATACCAACTAATC
2481  TCCAAATAAAACAAAACTCTTTGCTTTAATCTGGATTGTTGTGTCACAAATTAATGGGGA
```

Fig. 2. Nucleotide sequence and structure of the mouse *Sebox* gene and predicted amino acid sequence of Sebox protein. A portion of the *Sebox* genomic DNA that was sequenced is shown. Nucleotide residues shown in green represent two polyadenylation sites. Transcription initiation sites found in brain are shown in blue, and in ovary in red. The first ATG codon and two polyadenylation signals are shown in white letters on a black background. The homeodomain is shown in yellow. Arrowheads indicate splice sites. Introns are in lower case. An in-frame TGA termination codon upstream of the putative ATG initiation codon is underlined.

```
421   GGGCCACCCACTCTGAACGGCTTAAAAACCCTAAGGCCCAAGTCTGGTGTGGGACATAGCT
481   GGGGCATGGCTCTCTAGCTCCTCACCCTGTGTCAGCTCCATGGCCCTCTGTGGAGGCCA
  1                                          M  A  S  P  V  E  A
541   TCTCCAGgtgacctggtggggaaagggtaggtaggtcgaggaagtgtagcaaggtgggat
  8   S  P  G
601   ggggaagggtgaagagaggagggaataggggttgcagaggctgaccaagcctttgaggct
661   ggggtcttggtttgattcccatcctgacccacagGCTGTGCCAGTGGGTTGGGTCCCCAT
  11                                      C  A  S  G  L  G  P  H
721   CGGAGGAAGAGGACCACTTTTAGTGTTGGGCAGTTGCTGGAGCTGGAACGGGTATTTGCA
  19   R  R  K  R  T  T  F  S  V  G  Q  L  L  E  L  E  R  V  F  A
781   GCTAGGCCCTATCCTGACATCAGCACCCGTGAGCACCTGGCTCAGATAACTCACCTGCCA
  39   A  R  P  Y  P  D  I  S  T  R  E  H  L  A  Q  I  T  H  L  P
841   GAAGCCAAGATCCAGgtgagtcacaagagccaacttctgggctcaatacacagcctttc
  59   E  A  K  I  Q
901   tctctgggatgccctcaagttgttttctgtccagtgtgGtgcttgtgtgtgtgtgtgtgc
961   agtgcgcagacatgtgttggtgccaagatctatccctttatctttcctatcatgGTGTGGT
  64                                                         V  W  F
1021  TCCAGAACCGTCGAGCCAAGAGAATCAAGGACAGAAAGCCAGGAGCCCTAAACGCCAGGC
  67   Q  N  R  R  A  K  R  I  K  D  R  K  P  G  A  L  N  A  R  L
1081  TGGAGCTTCCCCCAAACTGTTCTCTTCCAGACACTCCCCAGCTATCTTGGGACCCTGGAA
  87   E  L  P  P  N  S  L  P  D  T  P  Q  L  S  W  D  P  G  T
1141  CACCAAGCCATCCTCTACACTCTACCAGTTCAGCTCAGCAGACTTCGGCATGCCCACCAC
  107  P  S  H  P  L  H  S  T  S  S  A  Q  Q  T  S  A  C  P  P  Q
1201  AGACCTCCTGCCTCAGGCCCCATTTTGGGTCCAGGGCAGAGTTGGTCAGGGGCTAAAGCTG
  127  T  S  C  L  A  P  I  L  G  P  G  Q  S  W  S  G  A  K  A  A
1261  CAGCCCCATGGGGGACAAGTGAGGTGTCAGGAGTTCACTCTTTAGAGCAAATTGTTCCTC
  147  A  P  W  G  T  S  E  V  S  G  V  H  S  L  E  Q  I  V  P
1321  AGACTTCACTGGGCAACCTGTCTGACCTTATCTATACCTCAGCCATCGTCACCAATGTAG
  167  T  S  L  G  N  L  S  D  L  I  Y  T  S  A  I  V  T  N  V  D
1381  ACCACTCCTAACTTAGGTGTAGAGCTTTTAAAACCTTGGTTTTGGCTTTATAGCCAGCCCA
  187  H  S  *  188
```

Fig. 3. Nucleotide sequence of rat *Sebox* genomic DNA and predicted amino acid sequence. A portion of the nucleotide sequence of rat *Sebox* genomic DNA is shown. The homeobox and the predicted homeodomain protein are enclosed in shaded boxes. Arrowheads indicate sites for RNA splicing. Introns are in lowercase. An in-frame TAA termination codon upstream of the putative first ATG codon is underlined. Four polymorphic sites (940G/A, 1073C/T, 1271G/T, and 1282A/G) are shown in bold. The nucleotide sequence and predicted amino acid sequence shown correspond to haplotype A.

UTR of *Sebox* cDNA. However, only the second polyadenylation site (nucleotide residue 2507) is extensively used, both in late-stage embryos and adult brain, as demonstrated by RNase protection assays with a Sebox antisense RNA probe spanning the region containing the two polyadenylation sites (data not shown). Therefore, the most frequently used polyadenylation signal is AATAAA (nucleotide residues 2485–2490), whereas ATTAAA (nucleotide residues 2321–2326) is used less frequently, as expected (10). The 5′ end of *Sebox* cDNA was cloned by RLM-RACE; therefore, the sequence of the 5′ region of *Sebox* cDNA and the transcriptional start sites were determined at the same time (5). The promoter region of the *Sebox* gene lacks an obvious TATA box (see Fig. 2), which often results in the use of multiple transcriptional initiation sites by RNA polymerase II (11). In Fig. 2 are shown the RNA initiation sites of *Sebox* transcripts isolated from adult ovary and brain poly(A)⁺ RNA. Four transcriptional initiation sites within a 300-nt region were found in *Sebox* mRNA from adult brain, which is in agreement with the diffuse appearance of two bands of RNA detected by Northern analysis (Fig. 1C). *Sebox* transcripts from ovary also were heterogeneous at the 5′ end, although different nucleotide residues are used in the two tissues as transcriptional start sites. The predicted first methionine of mouse Sebox protein is encoded by the ATG codon at position 978, which occurs in an adequate context for translation initiation (11). The optimal context for initiation of translation in vertebrates is

GCCRCCatgG (11). Within this consensus sequence, two nucleotide residues have the strongest effect: the G residue following the ATG codon (position +4) and a purine, preferably A, three nucleotides upstream (position −3). Translation initiation sites in vertebrate mRNAs nearly always conform to the consensus sequence in at least one of these two key positions (11). The nucleotide sequence flanking the ATG codon at position 978 of Sebox genomic DNA (Fig. 2) conforms to the Kozak consensus sequence at position +4, where a G residue is found. More upstream ATG codons that may represent sites for translation initiation are ruled out because of an in-frame stop codon (underlined TGA codon in Fig. 2) that would prevent the synthesis of a functional protein.

Three overlapping DNA fragments were amplified by PCR from rat genomic DNA, by using pairs of primers based on the nucleotide sequence of mouse *Sebox* genomic DNA. The nucleotide sequences of the amplified DNA fragments were assembled in a contig, 1,925 nucleotide residues in length, and submitted to GenBank. A portion of the consensus sequence (nucleotide residues 421-1440) and the predicted amino acid sequence of rat Sebox protein are shown in Fig. 3. Sequence analysis of 20 genomic DNA clones revealed four polymorphic sites: 940G/A, 1073C/T, 1271G/T, and 1282A/G (shown in white on a black background in Fig. 3). Interestingly, the polymorphic residues seem to be in linkage disequilibrium and segregate as two haplotypes, A and B. The sequence shown in Fig. 3 corresponds to the A haplotype, whereas the B haplotype is as follows: 940A, 1073T, 1271T, 1282G. Two of four polymorphic substitutions result in amino acid substitutions (149W→C and 153E→G), one is a silent mutation, and one (940G/A) is located within a predicted intron.

The *Sebox* gene was mapped to mouse chromosome 11, near the *Evi*2 locus. In 70 mice, no recombinants were found, indicating that the two genes are within 4.2 cM. This is a region of conserved synteny to a region of human chromosome 17 (12).

```
121   CTCCCTTCTTCCCCACCGAGACTTCTCTGCCTTGCTTGAGCTGGACTCTAAACCTCTCCA
181   CTCCTGAAGGCGGGGCCAGCCACCCTGAGCATGATAAAACCCAAATGGGGGGGTCTGGTG
  1                                   M  I  K  P  K  W  G  G  L  V
241   TGGGCACAGCCCTGGCATGGCCCTCTGGCACGGCCCTCTGGCACCTTACCCTTTGCCAGC
 11    W  A  Q  P  W  H  G  P  L  A  R  P  S  G  T  L  P  F  A  S
301   TCCATGCCCAGCCCTGTGGATGCATCCTCAGCCAGtgatctgatgggggaagtgtgggca
 31    S  M  P  S  P  V  D  A  S  S  A  D
361   gggtgagaacgcaggccatggtgaggatgaggacggttggagaggggtgggagtagggact
421   tcggggttgattgagggctttgaggctgggtggaggcttgcagggagtcccagcttga
481   ccccaccatggcctgcagACGGTGGCAGCGGGTTGGGTTCCCACCGGAGAAAGCGGACC
 43                      G  G  S  G  L  G  S  H  R  R  K  R  T
541   ACCTTCAGCAAAGGGCAGCTACTGGAGCTGGAGAGGGCGTTTGCAGCTGGCCCTACCCC
 56    T  F  S  K  G  Q  L  L  E  L  E  R  A  F  A  A  W  P  Y  P
601   AACATCAGCACCCATGAGCACTGGCTGGGTCACTTGCCTTCCTGAGGCCAAGGTACAG
 76    N  I  S  T  H  E  H  L  A  W  V  T  C  L  P  E  A  K  V  Q
661   gtgagctgtgaccccttctcacacagccaacatctaggcttggccacagcccttcgct
721   gggtagcctgcctcaggttgtttccctgtccctgggtggaggggatgctggggccaga
781   acagacccctcttctttcctagcagGTGTGGTTCCAGAAGCGCTGGGCCAAAATAATCA
 96                             V  W  F  Q  K  R  W  A  K  I  I  K
841   AGAACAGGAAGTCAGGAATTCTAAGCCCTGGGTCTGAGTGCCCCCAGAGCTCCTGTTCTC
108    N  R  K  S  G  I  L  S  P  G  S  E  C  P  Q  S  S  C  S  L
901   TTCCAGACACCCTCCAGCAGCCCTGGGATCCCCAAATGCCAGGCCAACCTCCACCCTCCA
128    P  D  T  L  Q  Q  P  W  D  P  Q  M  P  G  Q  P  P  P  S  S
961   GCGGCACACCTCAGCGGCACCTCAGTGTGTCGACATAGCTCCTGTCCAGCTCCTGGCTTGA
148    G  T  P  Q  R  T  S  V  C  R  H  S  S  C  P  A  P  G  L  S
1021  GTCCACGGCAGGGCTGGGAAGGGGCTAAAGCTGTAGCCCCATGGGGATCAGCTGGGGCTT
168    P  R  Q  G  W  E  G  A  K  A  V  A  P  W  G  S  A  G  A  S
1081  CAGAGGTCCACCCTTCTTTAGAGCGAGCTACTCCCCAGACTTCACTAGGCAGCCTGTCTG
188    E  V  H  P  S  L  E  R  A  T  P  Q  T  S  L  G  S  L  S  D
1141  ACCTCATCTATGCCTCGGCCATTGTCGTCAATGTGGACCACTCCTAGTCTTTGTCCAGAA
208    L  I  Y  A  S  A  I  V  V  N  V  D  H  S  *  221
```

Fig. 4. Nucleotide sequence of human *Sebox* genomic DNA and predicted amino acid sequence. A portion of the nucleotide sequence of human *Sebox* genomic DNA is shown. The homeobox and the homeodomain are enclosed in shaded boxes. Arrowheads indicate inferred sites for RNA splicing. Introns are in lowercase. An in-frame TAA termination codon upstream of the putative first ATG codon is underlined.

Accordingly, a human *SEBOX* gene was identified within a genomic sequence from human chromosome 17 (GenBank accession no. AC002094) by searching GenBank for *Sebox* homologs. We cloned, by PCR, two overlapping DNA fragments from human genomic DNA, spanning the *SEBOX* gene. A portion of human *SEBOX* DNA (nucleotide residues 121-1200) and the predicted amino acid sequence of human SEBOX protein are shown in Fig. 4. This sequence differs from the GenBank AC002094 sequence at two positions: a C residue at position 250, which is missing in GenBank AC002094, results in a frame-shift in the predicted amino acid sequence of SEBOX protein, and concomitantly a different amino terminal amino acid sequence; a nonsynonymous substitution at nucleotide residue 1156 (C–T) results in an L to S substitution in the C-terminal region of SEBOX protein. The inferred structure of the human *SEBOX* gene predicts three exons and thus resembles the rodent *Sebox* genes. The predicted intron/exon junctions, the termination codon, and the polyadenylation signals also are conserved. The human *SEBOX* gene encodes a putative protein that consists of 221 amino acid residues (Fig. 4), 31 residues longer than the mouse or rat counterparts. The additional amino acid residues are located in the N-terminal region and result from use of an upstream ATG codon (nucleotide residues 211–213). The nucleotide sequence surrounding the ATG codon is an adequate context, as it matches to the Kozak consensus sequence at one of the two key positions, i.e., the A residue at position −3. However, the possibility that the ATG at position 304 of human *SEBOX* mRNA is a site for initiation of protein synthesis cannot be ruled out.

In Fig. 5 are shown the amino acid sequence alignments of mouse, rat, and human Sebox proteins. Rat Sebox protein is 92% identical to mouse Sebox, with only two conservative amino acid substitutions within the homeodomain (97% homology). Rat Sebox lacks two serine residues in the C-terminal region that are present in the mouse protein (amino acid residues 92 and 160). Human and mouse Sebox proteins are 63% identical overall and 78% identical in the homeodomain. When considering conser-



```
mouse                                          MASPVEASPG.CASGLGPH   18
rat                                            ----------.--------   18
human   MIKPKWGGLVWAQPWHGPLARPSGTLPFASS-P---D--SADGG----S-   50

mouse   RRKRTTFSVGQLVELERVFAARPYPDISTREHLAQVTHLPEAKIQVWFQN   68
rat     ----------L---------------------I-----------------   68
human   -------K--L----A---W---N---H------W--C-----V-----K  100

mouse   RRAKRIKDRKPGALNSRLELPPNSCSLPDTPQLPWDPGTSSHPLHPTSSA  118
rat     ---------A----------S-----P-----S---                117
human   -W--I--N--S-I-SPGS-C-QS-------L-Q----QMPGQ-PPSSGTP  150

mouse   QYTSACPPQTSCLGPILGPGQSWSGAKVAAPWGTSGASGIHSSLEQIVPQ  168
rat     -Q-----------A-------------A--------EV--V-.-------   166
human   -R--V-.RHS--PA-G-S-R-G-E---AV----SA---EV-P---RAT--  199

mouse   TSLGNLSDLIYTSAIVTNVDHF*                             190
rat     ---------------------S*                             188
human   ----S------A----V----S*                             221
```

Fig. 5. Alignment of predicted amino acid sequences of mouse, rat, and human Sebox proteins. A dash represents an identical amino acid residue, a dot represents the absence of a residue. Conservative amino acid substitutions (according to Fig. 84 in ref. 23) are shown in bold; conservative amino acid replacement families are as follows: G, A, P, S, T; V, M, I, L; F, Y, W; D, E, N, Q; K, R, H; and C. Homeodomains are enclosed in the shaded box.

vative substitutions (shown in bold in Fig. 5), the homology increases to 78 and 87%, respectively. This degree of similarity is strikingly low. This divergent evolution of the *Sebox* gene in the two species might reflect the duplication of an ancestral gene, followed by rapid evolution of one of the genes (13). However, the significant homology and the location of both genes in syntenic chromosomal regions indicate that the two genes are true orthologs. In both genes, three exons are predicted. An intron between amino acids 46 and 47 of the homeodomain and intron–exon junctions, as well as polyadenylation signals, are conserved. One striking difference between the mouse and human predicted proteins is the substitution of Asn-51 and Arg-53 in the mouse homeodomain with Lys and Trp, respectively, in the human homolog. Sequence analyses of homeodomains identified so far show that Asn-51 is invariant, whereas position 53 is usually an Arg residue (2). Hence, the Asn-51 to Lys and Arg-53 to Trp substitutions raise the possibility that the human *SEBOX* gene is a pseudogene or, perhaps, raise doubt about the conservation of function of the human SEBOX protein.

Expression of *Sebox* in Maturing Oocytes and Preimplantation Embryos. The expression of *Sebox* mRNA in mouse adult ovary was analyzed further by *in situ* hybridization with a [^{33}P]-labeled RNA probe corresponding to the nucleotide sequence encoding the C-terminal portion of the Sebox protein. Developing ovarian follicles were staged by the size of the oocytes and by the number of layers of granulosa cells surrounding the oocyte, according to Pedersen and Peters (14). Expression of *Sebox* was detected in growing and full-grown oocytes, whereas somatic cells were not labeled (Fig. 6B). *Sebox* expression was not detected in small peripherally located resting oocytes in type 1 follicles (indicated as 1 in Fig. 6 *A* and *B*). *Sebox* RNA first appears at low levels in growing oocytes of monolayered follicles (type 3a/b), and the abundance of *Sebox* RNA increases as the follicles mature through the bilayered (type 4) and multilayered (type 5a/b) stages. High levels of *Sebox* RNA also were found in oocytes in early antral stage follicles (type 6). However, during late antral stages of follicle maturation, when meiosis resumes and oocyte transcription ceases, *Sebox* expression in oocytes decreases (not shown). *Sebox* mRNA was not detected in oocytes in atretic follicles, i.e., follicles that have stopped growing and will degenerate (not shown). These results demonstrate that *Sebox* is expressed in a developmentally regulated manner in the adult ovary and suggest a possible involvement of *Sebox* in the growth and acquisition of competence to resume meiosis in oocytes.

BIOCHEMISTRY
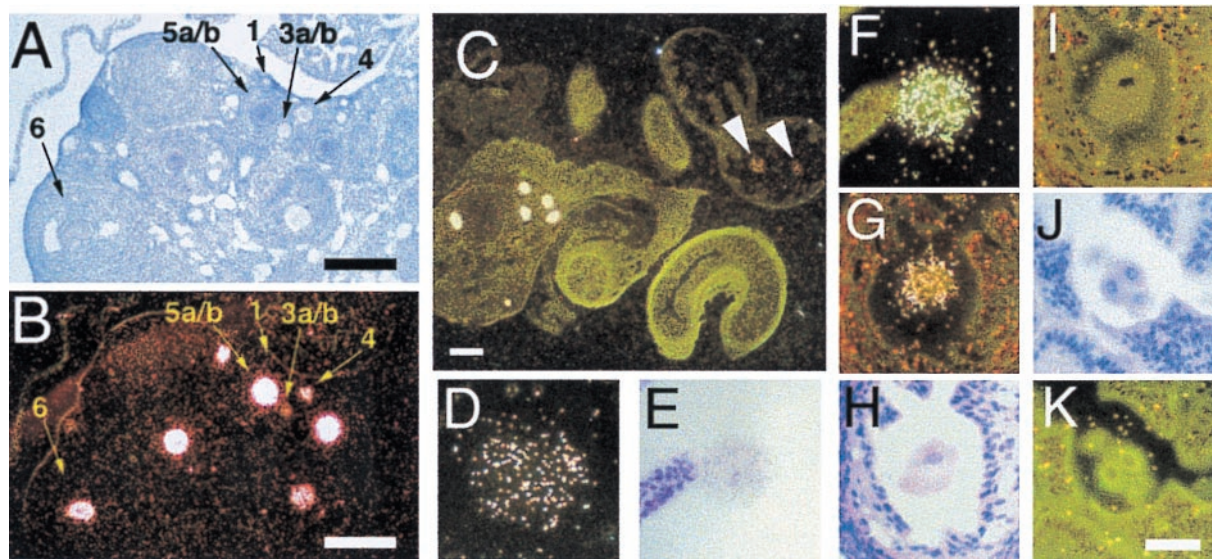
Fig. 6. *Sebox* gene expression in mouse oocytes, eggs, one-cell, and two-cell embryos. Shown are adult mouse ovaries (*A* and *B*), the ovary of a superovulated mice with the ampulla containing two eggs (*C*), unfertilized egg (*D*), fertilized egg (*E* and *F*), and two-cell (*G–I*) and four-cell (*J* and *K*) embryos. Sections were hybridized with either antisense (*B–D*, *F*, *G*, and *K*) or sense (*I*) *Sebox* RNA probes and photographed under brightfield (*A*, *E*, *H*, and *J*) or darkfield (*B–D*, *F*, *G*, *I*, and *K*) illumination. Sections in *A*, *E*, *H*, and *J* represent the brightfield images of *B*, *F*, *I*, and *K*, respectively. Sections in *G* and *I* are adjacent sections hybridized either with the antisense (*G*) or with the sense probe (*I*). No signal was detected with the (+) sense probe, demonstrating the specificity of hybridization. The stages of the follicles in *A* and *B* are indicated according to ref. 14: type 1, small oocyte with no follicle cells; type 3a/b, one complete ring of follicle cells surrounds a growing oocyte; type 4, two layers of follicle cells surround a growing oocyte; type 5a/b, many layers of follicle cells surround a large oocyte; type 6, large follicle with scattered areas containing fluid. Bars in *A*, *B*, and *C* represent 200 μm; bar in *K* represents 50 μm and applies to *D–K*.

Growing oocytes synthesize and accumulate RNAs, proteins, and organelles that constitute the maternal contribution to early development; then, during the antral stage of follicle maturation, oocytes initiate the degradation of maternal mRNAs, which is essentially complete by the end of the two-cell stage (15–17). However, the synthesis of maternal proteins can persist beyond this time (18).

In situ hybridization experiments were performed to determine whether *Sebox* is expressed in ovulated eggs, fertilized eggs, and early embryos. *Sebox* transcripts were detected in ovulated eggs (Fig. 6 *C* and *D*), fertilized eggs (Fig. 6 *E* and *F*), and two-cell stage embryos (Fig. 6*G*). Although *in situ* hybridization is a semiquantitative technique, analysis of about 50 sections for each stage consistently showed lower signals in eggs, fertilized eggs, and two-cell stage embryos, compared with the maximum intensity of signal detected in oocytes of
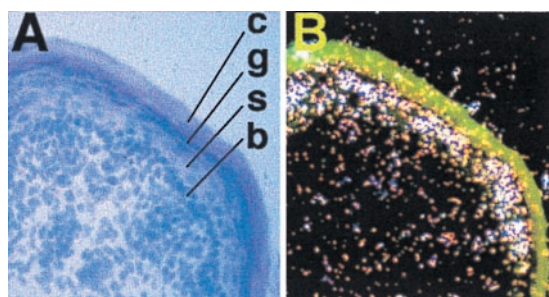
Fig. 7. *Sebox* gene expression in the epidermis of newborn mice. Shown is a representative section of skin from a newborn mouse, hybridized *in situ* with a mixture of two [^{33}P]-labeled antisense RNA probes encompassing the coding and noncoding regions of *Sebox* cDNA downstream of the homeobox (nucleotide residues 1491–1910 and 1914–2320, respectively, in Fig. 2). (*A*) Brightfield with eosin-hematoxylin staining. (*B*) Darkfield view of the same section. *Sebox* mRNA was detected mainly in the granular cell layer. c, cornified; g, granular; s, spinous; and b, basal layers.
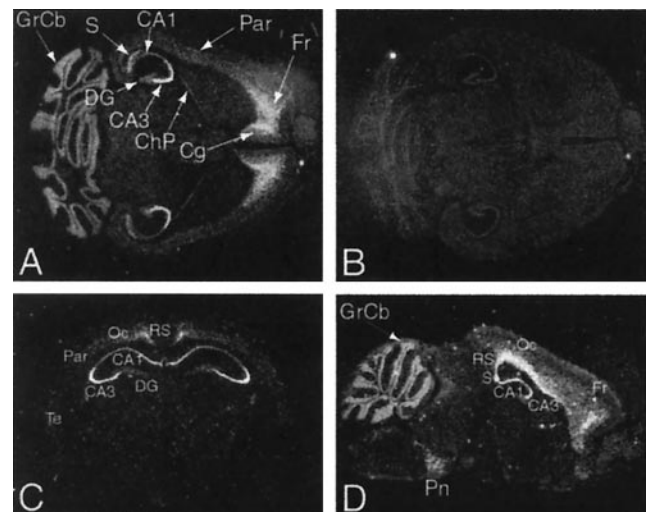
Fig. 8. *In situ* hybridization of adult mouse brain. Darkfield autoradiographs of representative horizontal (*A* and *B*), coronal (*C*), and parasagittal (*D*) sections hybridized with *Sebox* antisense (*A*, *C*, and *D*) or sense (*B*) [^{33}P]-labeled RNA probes. (*A*) Expression of *Sebox* RNA was detected in the cerebral cortex, CA areas of the hippocampus, subiculum, choroid plexus, and granular layer of the cerebellum. Within the hippocampus, intense labeling is present in the pyramidal cells of the CA3 region (*A* and *C*), whereas no *Sebox* RNA was detected in the granule cells of the dentate gyrus. In the cerebral cortex, *Sebox* is expressed at high levels in the frontal and cingulate cortex (*A*) and, more posteriorly, in the retrosplenial and occipital cortex (*C* and *D*). A lower level of *Sebox* mRNA was detected in the parietal and temporal cortex (*A* and *C*). In *D*, *Sebox* expression is shown in a continuous pattern that covers the CA regions of the hippocampus, the subiculum, and a deep layer of the cerebral cortex (retrosplenial, occipital, and frontal cortex). The pontine nuclei also are labeled (*D*). Abbreviations are as follows: CA1–3, CA regions of the hippocampus; ChP, choroid plexus; Cg, cingulate cortex; DG, dentate gyrus; Fr, frontal cortex; GrCb, granular layer of the cerebellum; Oc, occipital cortex; Par, parietal cortex; Pn, pontine nuclei; RS, retrosplenial cortex; S, subiculum; Te, temporal cortex.

Cinquanta *et al.*

multilayered follicles in the same tissue section. A representative section of an ovary and *ampulla* containing eggs is shown in Fig. 6*C*. The signals detected in ovulated eggs (indicated by arrowheads) are lower than the signals detected in oocytes within the same section. *Sebox* transcripts were not detected in four-cell embryos (Fig. 6 *J* and *K*).

Mouse zygotic transcription is initiated in the two-cell embryo (19, 20). We conclude that the *Sebox* gene is a maternally expressed gene. *Sebox* RNA detected in the two-cell embryo probably is a remnant of maternally expressed *Sebox* mRNA; however, the possibility of transient zygotic expression of the *Sebox* gene in the two-cell embryo is not ruled out.

Expression of *Sebox* in the Epidermis of Newborn Mice. Northern analysis revealed expression of the *Sebox* gene in the skin of newborn and adult mice. Expression of *Sebox* in skin of newborn mice was localized by *in situ* hybridization with a mixture of [^{33}P]-labeled RNA probes corresponding to the 3′ region of *Sebox* cDNA (see Fig. 7 legend). *Sebox* mRNA is specifically expressed in the differentiating suprabasal keratinocytes of the epidermis, whereas little or no *Sebox* mRNA was detected in the basal layer of the epidermis and the underlying dermis (Fig. 7). High levels of *Sebox* mRNA were detected in the granular layer of the epidermis, whereas little or no *Sebox* mRNA was detected in the spinous layer. *Sebox* gene expression was not detected in epidermal appendages, such as developing hair follicles and sweat glands, at least at this stage of development.

Vertebrate epidermis is a continually renewing tissue. Throughout life, stem cells in the basal layer undergo a multistage programmed differentiation as they migrate toward the surface of the organism (21). Several homeobox genes are expressed in fetal and adult epidermis, and the existence of additional regulatory genes has been postulated (21). Roles in the regulation of epidermal specific genes have been demonstrated in some cases. For instance, *Dlx3* mRNA is abundant in the granular layer, and there is evidence suggesting a role for this gene in the regulation of gene expression during late keratinocyte differentiation (22). Similarly, the expression of *Sebox* in the epidermis of newborn and adult mice raises the possibility that *Sebox* plays a role in keratinocyte differentiation.

Expression of *Sebox* in Adult Brain. Northern analysis showed that levels of *Sebox* mRNA gradually increase in brain during postnatal development. We then determined the distribution of *Sebox* mRNA in adult brain by using *in situ* hybridization. As shown in Fig. 8, *Sebox* mRNA was detected in the cerebral cortex, CA areas of the hippocampus, pontine nuclei, choroid plexus, and cerebellum. Within the Ammon's horn of the hippocampus, *Sebox* mRNA was detected in the pyramidal cells of the CA3 region at higher levels than in the CA1 and CA2 regions (Fig. 8 *A* and *C*). *Sebox* mRNA was not detected in the granule cells of the dentate gyrus (DG). In the cerebral cortex, relatively high levels of *Sebox* RNA were found in the frontal and cingulate cortex (Fig. 8*A*) and, more posteriorly, in the retrosplenial and occipital cortex (Fig. 8 *C* and *D*). Lower levels of *Sebox* mRNA were detected in the parietal and temporal cortex (Fig. 8 *A* and *C*). As shown in Fig. 8*D*, *Sebox* is expressed in a continuous pattern that extends from the CA regions of the hippocampus, the subiculum, to a deep layer of the cerebral cortex (retrosplenial, occipital and frontal cortex).

In conclusion, we have cloned mouse *Sebox* homeobox genomic DNA and cDNA, characterized its expression pattern, and isolated genomic clones for rat and human *Sebox* homologs. The expression pattern of *Sebox* is highly regulated in the mouse. *Sebox* RNA was detected in specific tissues and areas, from late-stage embryo onward, suggesting a role in the final differentiation or maintenance of function of specific cell types. Moreover, the expression of *Sebox* in growing and maturing oocytes, eggs, zygotes, and two-cell embryos suggests a role in the regulation of oocyte development and, perhaps, in the very earliest phases of embryonic development.

1. Gehring, W. J. (1994) in *Guidebook to the Homeobox Genes*, ed. Duboule, D. (Oxford Univ. Press, New York), pp. 3–10.
2. Burglin, T. R. (1994) in *Guidebook to the Homeobox Genes*, ed. Duboule, D. (Oxford Univ. Press, New York), pp. 27–71.
3. Rovescalli, A. C., Asoh, S. & Nirenberg, M. (1996) *Proc. Natl. Acad. Sci. USA* **93,** 10691–10696.
4. Valarché, I., Tissier-Seta, J.-P., Hirsch, M.-R., Martinez, S., Goridis, C. & Brunet, J.-F. (1993) *Development (Cambridge, U.K.)* **119,** 881–896.
5. Schaefer, B. C. (1995) *Anal. Biochem.* **227,** 255–273.
6. Adamson, M. C., Silver, J. & Kozak C. A. (1991) *Virology* **183,** 778–781.
7. Chomczynski, P. & Sacchi, N. (1987) *Anal. Biochem.* **162,** 156–159.
8. Hogan, B., Beddington, R., Costantini, F. & Lacy, E. (1994) *Manipulating the Mouse Embryo* (Cold Spring Harbor Lab. Press, Plainview, NY).
9. Mermod, N., O'Neill, E. A., Kelly, T. J. & Tjian, R. (1989) *Cell* **58,** 741–753.
10. Jacob, S. T., Terns, M. P., Hengst-Zhang, J. A. & Vulapalli, R. S. (1990) *Crit. Rev. Eukaryotic Gene Expression* **1,** 49–59.
11. Kozak, M. (1996) *Mamm. Genome* **7,** 563–574.
12. DeBry, R. W. & Seldin, M. F. (1996) *Genomics* **33,** 337–351.
13. De Robertis, E. M. (1994) in *Guidebook to the Homeobox Genes*, ed. Duboule, D. (Oxford Univ. Press, New York), pp. 13–23.
14. Pedersen, T. & Peters, H. (1968) *J. Reprod. Fert.* **17,** 555–557.
15. Bachvarova, R. & DeLeon, V. (1980) *Dev. Biol.* **74,** 1–8.
16. Bachvarova, R., DeLeon, V., Johnson, A., Kaplan, G. & Paynton, B. V. (1985) *Dev. Biol.* **108,** 325–331.
17. Bachvarova, R., Cohen, E. M., DeLeon, V., Tokunaga K., Sakiyama, S. & Paynton, B. V. (1989) *Development (Cambridge, U.K.)* **106,** 561–565.
18. Paynton, B. V., Rempel R. & Bachvarova, R. (1988) *Dev. Biol.* **129,** 304–314.
19. Braude, P., Pelham, H., Flach, G. & Lobatto, R. (1979) *Nature (London)* **282,** 102–105.
20. Flach, G., Johnson, M. H., Braude, P. R., Taylor, R. A. S. & Bolton, V. N. (1982) *EMBO J.* **1,** 681–686.
21. Sargent, T. D. & Morasso, M. I. (1999) in *Cell Lineage and Fate Determination*, ed. Moody, S. A. (Academic, San Diego), pp. 553–567.
20. Morasso, M. I., Markova, N. G. & Sargent, T. D. (1996) *J. Cell Biol.* **135,** 1879–1887.
23. Dayhoff, M. O., Schwartz, R. M. & Orcutt, B. C. (1978) in *Atlas of Protein Sequence and Structure*, ed. Dayhoff, M. O. (Natl. Biomed. Res. Found., Silver Spring, MD) Vol. 5, pp. 345–352.

BIOCHEMISTRY